Quantifying flow-ecology relationships with functional linear models

Ben Stewart-Koster¹,³, Julian D. Olden¹, Keith B. Gido²

¹ School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195, U.S.A. bensk@uw.edu

² Division of Biology, Kansas State University, Ackert Hall, Manhattan, Kansas, 66506, USA

³ Present address, Australian Rivers Institute, Griffith University, 170 Kessels Road, Nathan, Queensland 4111, Australia

Abstract Hydrologic metrics have been used widely to quantify flow-ecology relationships, however, there are several challenges associated with their use including the selection from a large number of available metrics and the limitation that metrics are a synthetic measure of a multi-dimensional flow regime. Using two case studies of fish species density and community composition, we illustrate the use of functional linear models to provide new insights into flow-ecology relationships and predict the expected impact of environmental flow scenarios, without relying on hydrologic metrics. The models identified statistically significant relationships to river flow over the 12 months prior to sampling (r² range 36–67%) and an environmental flow scenario that may enhance native species’ densities while controlling a non-native species. Hydrologic metrics continue to play an important role in ecohydrology and environmental flow management, however, functional linear models provide an approach that overcomes some of the limitations associated with their use.

Key words hydrologic metrics; environmental flows; ecohydrology; flow variation; river regulation; functional data analysis; dams

INTRODUCTION

Spatial and temporal variation in natural water flows is fundamental to the long-term sustainability and productivity of riverine ecosystems and their riparian areas (Poff et al., 1997). It is well recognized that river flow is the “master” variable that governs lotic ecosystems, driving strong ecological consequences at local to regional scales, and at time intervals ranging from days (ecological effects) to millennia (evolutionary effects) (Naiman et al., 2008). Past research has generated a wealth of knowledge aimed at establishing relationships between biotic assemblages and characteristics of the flow regime (e.g. Kennard et al., 2007; Stewart-Koster et al., 2011; Gido et al., 2013), particularly to support environmental flow efforts in human-altered systems (Arthington et al., 2006; Poff et al., 2010; Poff and Zimmerman, 2010). Researchers often approach such questions by describing key components of the hydrograph (i.e. magnitude,
frequency, timing, duration and rate of change of flows) using hydrologic metrics or indices based on a given time series of discharge data (Olden and Poff, 2003; Mathews and Richter, 2007; Gao et al., 2009; Kennard et al., 2010).

The long history of hydrologic metrics usage in both hydrology and ecology (Olden et al., 2012) has resulted in greater ecological understanding and more informed management of water resources (Richter et al., 2006). Pioneering work in this area focused on single metrics, or a small number that described typical hydrologic conditions, variability around averages, predictability of flow events or statistical characteristics of flow events such as skewness (Clausen and Biggs, 2000; Olden and Poff, 2003). With increasing interest and research on flow-ecology relationships came the proliferation of hydrologic metrics to simultaneously describe multiple characteristics of the flow regime (e.g., Poff and Ward, 1989; Richter et al., 1996; Puckridge et al., 1998). There are now literally hundreds of different metrics (Olden and Poff, 2003), many of which are associated with high statistical uncertainty (see Kennard et al., 2010 for multiple sources of uncertainty), from which researchers and managers must choose a conceptually and ecologically tractable subset.

Although the history of hydrologic metrics has yielded many important advances in river ecology and management, there continues to be several practical challenges associated with their use. The first is the task of selecting a subset of hydrologic metrics from a potentially bewildering array of possible choices. This can be aided by a formal redundancy analysis, however, it is generally understood that ecological knowledge should also be incorporated into such decisions (Olden and Poff, 2003). A second task is to determine environmental flow standards based on relationships between ecological response variables and potentially disparate hydrologic metrics (Arthington et al., 2006). With so many metrics available there is often little coordination among studies making synthesis and generalisations of flow-ecology relationships very difficult (Poff and Zimmerman, 2010).

A synthetic summary of a complex hydrograph provided by hydrologic metrics is beneficial in many circumstances, but this also creates a practical challenge. Hydrologic metrics are descriptors of a hydrograph, or some facet of it, rather than depicting the flow itself. The implication of this is that limited information is available about the mechanistic link between the response variable and the actual discharge reflected by the hydrograph; rather it is a relationship between the response variable and a specific characteristic of the discharge. Related to this, is the extreme example where two very different hydrographs may have the same or similar values for a set of hydrologic metrics over a given interval (e.g. Fig. 1). Although this is a conveniently chosen example, it does help to illustrate that a single value of a hydrologic metric could be derived from an almost infinite number of differently shaped hydrographs. In such cases, using hydrologic metrics without judicious selection may lead to difficulties in establishing flow-ecology relationships and could compromise the success of environmental flow management. Consequently it is worthwhile exploring alternative approaches to relying solely on hydrologic metrics for river research and management.

Recent statistical advances can provide a more holistic characterization of the flow regime in an attempt to develop meaningful hydrologic descriptors. Spectral analysis techniques such as Fourier and wavelet analysis offer quantitative approaches that incorporate a more continuous hydrologic perspective to quantify meaningful descriptors of the flow regime (see Sabo and Post, 2008; and Steel and Lange, 2007 for an example with thermal regimes). These methods aim to decompose the mathematical functions that underlie the observed time series for use in subsequent statistical analyses (Priestley, 1991). Building on these analytical approaches
is the field of functional data analysis, which has provided a framework to incorporate such spectral decompositions of functional data into statistical models (Ramsay and Silverman, 2006). Functional data can have many forms but typically consist of continuous, or effectively continuous observations of a process over space or time that are derived from mathematical functions (Ramsay and Silverman, 2002), such as temperature or river discharge (e.g., Ainsworth et al., 2011). Quantifying and understanding the functional data series may itself be the goal of the analysis or they may be used as predictor variables in a functional linear model (Ramsay and Silverman, 2002; Müller and Stadtmüller, 2005). In such models, the actual time series, in this case the hydrograph, becomes the predictor variable in a linear model as opposed to a set of hydrologic metrics that aim to describe it. These functional methodologies may provide a valuable tool to overcome some, but not necessarily all, of the challenges associated with the use of hydrologic metrics in ecohydrology. In addition to this they may be useful in identifying flow-ecology relationships for hitherto unstudied species.

In this study we demonstrate the use of functional linear models to quantify flow-ecology relationships in riverine ecosystems. We use two case studies to illustrate the flexibility of the statistical approach; the first consists of data characterising spatial variation in the frequency of species occurrence (summarized into life history strategies) for a single year from 32 streams across North America and the second comprises time series data on fish densities in a single river for 18 years. These case studies were selected to represent the most common data types being analysed in the literature to establish flow-ecology relationships and inform the science of environmental flows.

METHODS

Functional linear models
Functional linear models can have several forms including the case where the predictor variable is a time series and the response variable is a scalar (Ramsay and Silverman, 2002). In the context of ecohydrology, such a model would relate the observed value of a response variable at a single point in time (or space) to the hydrograph over a specified time interval. That is each observation consists of a time series of flow and a single response variable. Thus hydrologic metrics are replaced as predictor variables by the time series of flow from which they would be calculated. To use the time series of flow in this manner requires a functional transformation of the hydrograph, such as a Fourier or B-spline transformation. From this, the integral of the curve can be estimated and then used as the functional predictor variable (Ramsay and Silverman, 2006). In this way the predictor variable in the statistical model is the actual hydrograph. The mathematical form of the model is that of an integral equation:

$$y_i = \beta_0 + \int x_i(t)\beta_1(t)dt + \epsilon_i$$

where, $y_i$ is the $i$th value of the response variable, $\beta_0$ is the intercept term of the regression, $x_i(t)$ is the value of the hydrograph for the $i$th observation at time $t$, $\beta_1(t)$ is the regression coefficient describing the relationship between the response variable $y_i$ and the predictor variable $x_i$, at time $t$ and $\epsilon_i$ is the residual of the $i$th observation in the model. This is the general form of the functional linear model, however, the residuals may be assumed to follow
a normal distribution or certain exponential distributions (e.g. Poisson, Müller and Stadtmüller, 2005). In this form of a functional linear model, the regression coefficient is itself a mathematical function that describes the relationship between the response variable and the predictor variable at each time $t$. In the context of flow-ecology, this model quantifies the relationship between the observed values of the response variable (which are scalars) and each day of the time series of flow. This effectively quantifies the importance of the timing, duration and magnitude of key flow events. Additional scalar predictor variables i.e. those used in standard regression modelling, such as habitat descriptors, can also be included in addition to the functional predictor variable (Ramsay and Silverman, 2006). An advantage of functional linear models is that through the use of basis functions in the modelling process, the regression coefficient can be estimated using the standard linear regression equation (Ramsay and Silverman, 2006). A brief explanation follows below.

A key step in fitting a functional linear model is to define the functional predictor variable, in this case the hydrograph, in an appropriate form for the model. This step essentially converts the observed flow data into a mathematical function that describes the time series, which can then be used in the regression. The function itself is not interpreted; it is only required to convert the observed time series into a form that can be used as a predictor variable in the functional linear model. This can be done by either transforming the hydrograph into a Fourier series or a B-spline series (Ramsay and Silverman, 2006). Fourier series are a decomposition of a periodic series into a sum of sine and cosine waves and are typically used to quantify the frequencies of periodic patterns in time series such as daily temperatures (Sabo and Post, 2008). While seasonality due to rainfall and snowmelt clearly exist in long term hydrologic data, periodic patterns are likely to be absent across relatively short intervals (i.e. one year or less). In such cases a B-spline curve can be used to define the hydrograph as a functional predictor variable (e.g. Ainsworth et al., 2011). B-spline curves, which are a series of joined polynomial functions, offer great flexibility in quantifying the functional form of time series data (Ramsay and Silverman, 2006). Converting the time series to a set of basis functions creates a linear combination of basis functions with a set of its own coefficients:

$$x_i(t) = \sum_{k=1}^{K_x} c_{ik} \psi_k(t),$$

which in matrix format is $x_i(t) = C\psi(t)$, where $C$ is the $N$ by $K_x$ coefficient matrix for the vector of $K$ basis functions, $\psi$.

The next step in model development is to define the form of the functional regression coefficient, $\beta_1(t)$ the shape of which is estimated in the model fitting process (Ramsay and Silverman, 2006). Like the predictor variable, this can take the form of a Fourier series or a B-spline curve thereby producing a linear combination of basis functions:

$$\beta(t) = \sum_{k=1}^{K_\beta} b_k \theta_k(t).$$

In matrix form this is, $\beta = \theta'b$, where $\theta$ is a vector of basis functions and $b$ is the vector of $K_\beta$ coefficients.

Having defined the functional predictor variable and the functional regression coefficient, the model can now be defined as:
In matrix form this is \( y_i = \int C\psi(t)\theta(s)'b dt \)

In matrix form this is \( y_i = CJ_{\psi\theta}b \), where \( J_{\psi\theta} \) is a \( K \times K \) matrix defined by

\[ \int \psi(t)\theta'(s) dt \]

The matrix form of the model can be further simplified so that \( K+1 \) vector comprised of an intercept \( \beta_0 \) and all \( k \) coefficients = \( \zeta \), and the intercept and coefficient matrix associated with the basis functions of the functional predictor = \( Z \). Then the model can be written out as

\[ \hat{y} = Z\zeta \]

which can be solved using the standard linear regression equation:

\[ \zeta = (Z'Z)^{-1}Z'y \]

An alternative to using basis functions would be to fit a standard multiple regression model using each observed daily flow value separately for the entire time period of interest, thereby using \( t \) predictor variables. However, this would very likely result in an overfit model with little predictive value with serious problems of multicollinearity. A functional linear model provides an efficient alternative to summarise the time series into a smaller number of basis functions, however, the potential for overfitting still exists with functional linear models. Therefore, the challenge is to use a functional regression coefficient that is sufficiently complex to capture relationships between the response variable and the time series of the predictor, yet is simple enough to interpret and not result in an overfit model (Ramsay et al., 2010; Ainsworth et al., 2011). As is common in most statistical models, this amounts to finding the trade-off between detecting signal from noise. One way to achieve this is to start with a large number of basis functions and use a smoothing parameter to penalise the “roughness”, or amount of curvature of the regression coefficient (Ramsay and Silverman, 2006). Such penalty terms are common in information theoretic approaches to model selection, such as AIC and BIC. The smoothing parameter is used as a coefficient for the “roughness” of the predictor function in an estimate of penalised sums of squares (PENSSE):

\[
PENSSE(\alpha_0, \beta) = \sum[y_i - \alpha_0 - \int x(t)\beta(t)dt]^2 + \lambda \int [L\beta(t)]^2 dt
\]

Where \( \int [L\beta(t)]^2 dt \) is a linear differential operator used to define the roughness of the functional regression coefficient, and \( \lambda \) is the smoothing parameter that penalises it (Ramsay et al., 2010). Minimising the PENSSE provides a functional regression coefficient for a given value of \( \lambda \) with degrees of freedom equal to the trace of the hat matrix from the regression. Selecting the value of \( \lambda \) can be done via cross-validation tests of multiple possible values providing an automatic selection of the smoothed regression coefficient for the best fitting model (Ramsay and Silverman, 2006). This results in a functional regression coefficient that minimises the penalised sum of squares and is directly interpretable as the relationship between the predictor variable over the time series and the response variable. The shape of the fitted functional regression coefficient can be illustrated graphically as a curve over the time interval of interest. In the context of flow-ecology relationships, the response variable shows a positive association to higher flows at times of the interval that the curve is above the zero line. Conversely, at times of the interval that the curve is below the zero line, the response variable shows a negative association with higher flows. A significant relationship to the predictor...
variable at any given time is evident when the 95% confidence intervals around the functional regression coefficient do not overlap zero (Ainsworth et al., 2011). Approximate point-wise confidence intervals are estimated as two times the standard error of the regression coefficient at time $t$ (Ramsay et al., 2010). While currently not available in freely downloadable software, it is also possible to derive joint confidence regions to enable simultaneous inference thereby better accounting for multiple testing (Ainsworth et al., 2011).

Permutation tests are used to test the significance of the functional linear model because it is difficult to derive a theoretical null distribution (Ramsay et al., 2010). The process follows a standard permutation testing approach where an $F$-statistic for the initial model fit is compared against a null distribution obtained by calculating the same statistic for a large number of model fits from random permutations of the response variable (Manly, 2007). In the case of functional linear models, the $F$-statistic is a comparison of the residual variance and the predicted variance for each permuted model fit (Ramsay et al., 2010). Comparing the original $F$-statistic with the permuted null distribution provides an estimate of a $p$-value for the statistical significance of the model fit.

For a full explanation of functional data analysis, including functional linear models, we refer the reader to Ramsay and Silverman (2006) and Ramsay et al. (2010). Below we demonstrate the use of functional linear models with two case studies where the predictor variable is a time series of flow for each observation and the response variable is a single value, or scalar, for each observation.

**Case study #1 Flow determinants of fish life history composition in Midwestern USA**
The first case study comprised a dataset of fish species occurrences from one-time community surveys at 32 free flowing streams with minimal human impact in the midwestern region of the United States (Mims and Olden, 2012). All streams are located within five contiguous freshwater ecoregions (Abell et al., 2008); Upper Missouri, Middle Missouri, English-Winnipeg Lakes, Upper Mississippi and Laurentian Great Lakes (Fig. 2). These streams are arrayed across a range of elevations and experience broadly similar continental climates (Abell et al., 2008). The study streams are predominantly perennial with peak flows from March to May each year (spring) and low flows occurring from August to October (late summer). In all cases, sampling occurred during summer months, defined for our purposes as 1 July. The dataset used here is a subset of that used in Mims and Olden (2012) and is fully described therein.

At each sampling site, species were classified as opportunistic, periodic or equilibrium according to the life history model of Winemiller and Rose (1992). These classifications were determined based on each species’ relative investment in reproductive parameters of fecundity, onset and duration of reproductive period and parental care (following Mims and Olden, 2012). Periodic strategists are those that are large bodied, long lived, and late-maturing species, opportunistic strategists are small bodied, short lived, early-maturing species and equilibrium strategists are medium bodied species with moderate age to maturation and generally provide parental care (Winemiller and Rose, 1992). Each of these life history strategies has different hydrologic preferences with periodic strategists favouring highly seasonal and periodically suitable hydrology while opportunistic strategists favour highly variable and unpredictable hydrologic environments and equilibrium strategists tend to favour habitats with relatively stable hydrology (Olden et al., 2006; Tedesco et al., 2008; Olden and Kennard, 2010). At each site, we calculated the relative proportion of species in each life history strategy according to species presence-absence.
Equilibrium species are known to prefer stable flows, however, flow stability was confounded with flow magnitude across the sampling reaches in this case study. Lower elevation reaches with more stable discharge also had generally higher flows. To prevent this confounding from influencing model fit, we standardized daily discharge by drainage area and related the relative proportion of each life history strategy to daily runoff (discharge/drainage area).

Case study #2. Flow determinants of fish species densities in the San Juan River
The second case study consists of a single sample of fish abundance collected each year over 18 years from secondary channels in a single 138 km long reach of the San Juan River in New Mexico and Utah from 1993–2010 (see Gido and Propst, 2012). Being a semi-arid river, the hydrology of the San Juan River was historically highly seasonal and also highly variable, both inter and intra-annually, with peak flows occurring during spring snowmelt. The sampling reach is in a segment of the river that is partly regulated by the Navajo Dam, which is managed to mimic the natural flow regime where possible (Propst and Gido, 2004). To this end, there is approximately 370,000 ML of water available per year for environmental flows (Holden, 1999). This allocation has been used to deliver flows that augment the spring flows to improve conditions for endangered species downstream (Holden, 1999). In some years such as 2008 (Fig. 3b) releases may continue into early summer (Bliesner et al., 2009). Despite the mimicry of the natural flow regime by dam operators, the biggest change since impoundment has been the reduction in hydrologic variability as well as reduced mean spring flows (Propst and Gido, 2004).

For this case study, we focused on three species with known relationships to temporal variation in discharge, the native species flannelmouth sucker *Catostomus latipinnis* and speckled dace *Rhinichthys osculus* as well as the non-native species red shiner *Cyprinella lutrensis*. Flannelmouth sucker is considered a periodic species (Olden et al., 2006) with spawning in the San Juan River beginning as early as March and concluding in July (Gido and Propst, 1999), but potentially continuing until October in other rivers (Minckley and Marsh, 2009). It has shown positive associations with increased flows in spring (Gido and Propst, 2012) while high flows in summer have been associated with improved fish condition (Paukert and Rogers, 2004). Speckled dace is a riffle dwelling species that generally prefers shallow, higher velocity habitats (Minckley and Marsh, 2009). It is considered an equilibrium species (Olden et al., 2006) with spawning occurring in spring with a possible second spawning event in the late summer (John, 1963). Speckled dace has shown positive associations with high flows in spring (Gido and Propst, 2012). The non-native red shiner is a hardy species that is very tolerant to highly variable flows (Moore and Thorp, 2008) and is recognized as an opportunistic species (Zeug and Winemiller, 2007). It has a long potential spawning season that can run from April–September (Farringer et al., 1979), but in the San Juan River, spawning is often delayed in years with high spring discharge (Franssen et al., 2005). Moreover, extended low flow events during the latter part of the spawning season are positively associated with increased densities (Propst and Gido, 2004; Gido and Propst, 2012).

The data collection for this case study occurred between 15 September and 15 October each year and standardized sampling methods were used to ensure each habitat type (e.g. riffle, run or pool) within the reach was sampled according to its proportional area (Propst and Gido, 2004). Species abundance was represented by densities and based on the number of individuals caught within the area of habitat sampled (count/m$^{-2}$). Standardizing by habitat area sampled
facilitated inter-annual comparisons for this analysis. Subsequently, we related the log transformed single observed density for each species in the autumn of each year to the log transformed daily discharge for the full year prior to sampling. Transforming the discharge data was required to reduce the skewness in the predictor variable and facilitate parameter estimation. See Propst and Gido (2004) and Gido and Propst (2012) for more details regarding the sampling methods and results from earlier analyses using of hydrologic metrics.

Having developed the functional linear models for the three species we then illustrate the utility of the approach for managing flow regimes by predicting species responses to hypothetical hydrographs. We derived five hypothetical hydrographs by “releasing” the water available for environmental flows at different times of the year. To do this we subtracted approximately 370,000 ML from the spring flow release in the 2009 water year and added this water to the hydrograph at different times of the year, resulting in five hypothetical environmental flow scenarios (Fig. 4). These scenarios consisted of a single flow release for 30, 60 or 90 days which were achieved by “releasing” 12335, 6167 or 4070 ML per day respectively. This equates to mean daily flow rates of 143 m$^{-3}$s$^{-1}$, 71 m$^{-3}$s$^{-1}$ and 48 m$^{-3}$s$^{-1}$. We subsequently applied the fitted functional linear models to predict the density of each species expected under these hypothetical hydrographs.

Each of these hypothetical hydrographs, used to demonstrate the predictive capacity of the approach, falls within the bounds of the observed hydrographs during the 18 years of data used to develop the models.

**Model parameterisation**

We developed the functional linear models with normally distributed errors in the R statistical environment (R Development Core Team, 2012) using the contributed package, fda (Ramsay et al., 2011). A graphical residual analysis for each model showed no systematic pattern in the residuals (results not shown). For both case studies we followed Ainsworth et al. (2011) by using B-splines with knots placed every five days to define the functional predictor variable for each observation. Two functional predictor variables are shown in Fig. 3, to illustrate the output from the mathematical function that is used as the predictor variable. We used the same basis to define the functional regression coefficient. We tested 29 different values for the smoothing parameter, $\lambda (10^1, 10^{1.5}, 10^2, 10^{2.5} \ldots 10^{14.5}, 10^{15})$, and identified the best fitting, smoothed regression coefficient via generalised cross-validation (Ramsay et al., 2010). Ainsworth et al., (2011) provide a useful discussion on selection of smoothing parameters with respect to generalised cross-validation.

The process of estimating predicted values under the model given the environmental flow scenarios follows the same approach as an ordinary linear model where the new predictor variables are applied to the coefficients in the regression equation. All that is required is to convert the hypothetical hydrographs into functional predictor variables using the same basis (in this case, a B-spline curve). This is easily achieved using the fda package in R (Ramsay et al., 2011).

As noted above, functional linear models are flexible enough to accommodate functional and scalar predictor variables in the same regression. This may be particularly useful to capture relationships between fish and hydrologic metrics describing flow variability or counts of certain types of flow days which would not otherwise be well quantified by a functional predictor. To demonstrate this we also fit a hybrid model for red shiner in Case study # 2, using a functional predictor for one part of the year and two hydrologic metrics for other times of the year. Given
the relationships between this species and flow, described in the literature and outlined above, we used spring flow as a functional predictor variable (1 March–31 May), again using a B-spline with knots placed every five days. Based on previous research (Propst and Gido, 2004; Gido and Propst, 2012), we used the number of days with flow < 14 m$^{-3}$ s$^{-1}$ between 1 June and 31 August (LFS) and the number days with flows above the 75th percentile for the period 1 October – 31 December (HFA) as hydrologic metrics in the model. When using additional scalar predictors in this way the interpretation of the regression coefficients is the same as any multiple regression model; each coefficient quantifies the effect of each predictor variable having controlled for all others in the model (Ramsay et al., 2006). Statistical significance of the regression coefficient(s) across the time interval of each model was assessed using pointwise confidence intervals (Ramsay et al., 2010), however, using the more stringent joint confident regions which account for multiple testing may produce slightly different results (Ainsworth et al., 2011).

RESULTS

Case study #1 Flow determinants of fish life history composition in Midwestern USA
As predicted by ecological theory, the magnitude and direction of the relationship between runoff and the proportion of each life history strategy varied among the life history strategies. The models explained a considerable amount of variation for each strategy (range: 36–62%), where permutation testing suggested each model was at least marginally significant (p-values shown on Fig. 5a–c). The functional regression coefficients for periodic strategists (p=0.03, 5.8 degrees of freedom) and equilibrium strategists (p=0.07, 5.8 degrees of freedom) showed very similar relationships to runoff in both magnitude and direction. This is despite these two groups showing little correlation in their relative richness across sites ($r = 0.18$). The two coefficients indicate significant negative associations with higher runoff in March and April and a positive association to higher runoff in May and June. A key difference between the regression coefficients of these two groups was the marginally significant positive relationship to runoff for equilibrium strategists in December–January during which periodic strategists showed no relationship. By contrast, the functional coefficient for opportunistic strategists (p<0.001, 6.6 degrees of freedom) was almost opposite to that of equilibrium strategists (Fig. 5b) with a significant positive association with runoff in March and April but a negative association in May and June. The positive association between equilibrium strategies and runoff in July and August shown is mirrored by corresponding negative associations for opportunistic strategists.

Case study #2. Flow determinants of fish species densities in the San Juan River
The functional linear models identified significant relationships to flow at different times of the year for each species in the San Juan River (Fig. 6a–c). The models for speckled dace (p=0.03, 3.8 degrees of freedom) and flannelmouth sucker (p=0.13, 4.2 degrees of freedom) both explained approximately 40–45% of the temporal variation in the species’ densities though the model for flannelmouth sucker was not a significant fit. The model for red shiner was also not a significant fit and only explained 14% of the variation. Speckled dace and flannelmouth sucker showed negative relationships to higher flows in October and November though only the relationship for speckled dace was significant at this time (Fig. 6a and b). The regression coefficients for these two species diverged in February and March when flannelmouth sucker showed a positive relationship to flows. By contrast, the model for speckled dace showed a more
gradual change to positive associations with flows, which peaked in May and June. Flows in late summer were not significantly related to the density of either species.

The hybrid model for red shiner, which included a three-month functional predictor and two hydrologic metrics, was a considerably better fit than the first model that used an entire year’s flow as the functional predictor variable \((p<0.001, \ r^2 = 0.67; \text{Fig. 7})\). Interestingly, by including the two hydrologic metrics along with the shorter interval functional coefficient, the model identified a significant positive relationship to higher flows in April–May (Fig. 7a) that was not evident in the first model (Fig. 6c). In addition, the model identified a positive association with higher numbers of low flow days in summer, LFS, and a negative association with higher numbers of high flow days in the late autumn period of October–December, HFA (Fig. 7b).

As expected, the results from the environmental flow scenarios were consistent with the shapes of the fitted regression coefficients. Predicted density for the two native species was generally higher than the respective 18-year average when the timing of the hypothetical release coincided with a positive relationship to flow for that species. For example, flannelmouth sucker showed the highest predicted density when the environmental flow was “released” in early spring for 60 days (Table 1), the time when this species showed a significant positive relationship to higher observed flows (Fig. 6b). Equally, the density of all species was much lower than average when the flow was released in late autumn (Table 1). Speckled dace and flannelmouth sucker both showed a negative relationship to higher flows at this time (Fig. 6a and 6b) while red shiner showed a negative relationship to the number of days above the 75th percentile of flow (HFA, Fig. 7). The 60-day early spring flow showed the greatest capacity to enhance densities of the two native species while somewhat depressing abundances of the non-native, red shiner.

**DISCUSSION**

In the present study we outlined an alternative approach to fitting flow-ecology relationships made available by advances in the area of functional data analysis. We sought to apply this approach to facilitate an accurate description of the hydrograph over a given time interval for use in statistical models. The models provided a good fit to the data in each case study and identified periods of each water year where there was a significant relationship to river flow. The regression coefficients in our models quantify relationships between the response variables and actual river flow and there is limited capacity to quantify relationships to flow variability with a functional predictor variable. However, the models can provide managers with a specific understanding of the timing, magnitude and duration of ecologically important flow events to replicate when setting environmental flow standards. In addition, functional models may provide a valuable tool for ecologists seeking to explore the flow-ecology relationships of poorly understood species.

**Flow-ecology relationships: the case studies**

Functional linear models provided new insight into flow-ecology relationships in both case studies. Previous analyses of relationships between fish life history strategies and hydrology
have combined data from broad geographic regions, and subsequently relied on hydrologic descriptors of variability and predictability for analysis (e.g., Olden and Kennard, 2010; Mims and Olden, 2012). These studies supported the expectation that periodic strategists are associated with streams with highly predictable, seasonal flows, opportunistic strategists are associated with streams with highly variable flows, and that equilibrium strategists would show variable responses to these flow characteristics. In the present study we restricted the analysis to a specific region and used daily runoff (flow divided by catchment area) as the predictor variable to quantify associations with the timing, magnitude and duration of annual discharge events.

Periodic strategists demonstrate strong reliance on seasonally predictable changes such as high flow pulses at specific times of the year (Winemiller and Rose, 1992). This was reflected in the regression coefficient for these species, as the significant relationship to elevated runoff was confined to the relatively predictable spring/summer period. Opportunistic strategists showed significant relationships to runoff at different times of the year including a positive association to higher flows in spring. This is consistent with the ecological expectation that rapidly maturing opportunistic species can take advantage of quickly changing flows that are associated with spring thaw and rain events (Winemiller and Rose, 1992). However, the significant negative association with runoff from May to September highlights that protracted disturbance events in the form of high flows across the entire spawning season have a negative impact on reproductive success of these species (Cooke et al., 2005; Roberts et al., 2006). Similar to opportunistic strategists, equilibrium strategists show significant associations to runoff at several times of the year, including a positive association with runoff in early winter. Such a relationship may reflect the gain from extended access to near-shore and floodplain habitat and associated warmer temperatures in the lead up to overwintering that is beneficial for species with high parental investment. For example, overwinter survival of juvenile smallmouth bass is enhanced by warmer temperatures and higher winter flows provided by groundwater fed springs (Peterson and Rabeni, 1996; Brewer, 2012).

In the second case study, functional linear models were successful in identifying flow-ecology relationships over time in the San Juan River. The model for flannelmouth sucker revealed a positive association between early spring flows and higher densities; likely reflecting their use of backwater habitats for spawning, which are accessible following elevated spring discharge (Bliesner et al., 1999). Thus, higher flows in spring would potentially result in a corresponding higher density of young-of-year suckers in the early fall when the data were collected. Though not a significant model fit according to generalized cross validation, the explanatory power of this model ($r^2 = 0.43$) is comparable to that of a multiple regression model fit with the same data but using hydrologic metrics ($r^2 = 0.44$: Gido and Propst, 2012). Similarly for speckled dace, a riffle-dwelling species that spawns in spring and summer, higher flows throughout this period were positively associated with higher densities likely related to spawning success (Bliesner et al., 1999; Gido and Propst, 2012). The explanatory power of this model ($r^2 = 0.42$) is considerably higher than that of the multiple regression model using hydrologic metrics ($r^2 = 0.29$: Gido and Propst, 2012). The first model for red shiner, using only a functional predictor variable for the entire year failed to identify a significant relationship to river flow. This is perhaps not surprising given red shiner is a habitat generalist that is tolerant of a wide variety of hydrologic conditions (Minckley and Marsh, 2009).

The second and better fitting model for red shiner was more complicated and thereby able to quantify relationships to very specific characteristics of the hydrograph as well as actual discharge. The model quantified the likely negative impact of higher numbers of high-flow days
in autumn as well as the positive effect of spring flows and higher numbers of low-flow days in summer. These relationships are probably related to the spawning strategy of this species and the impact of the timing of flow events on spawning success. High flows in spring, which have a positive association with red shiner density according to the functional regression, provide access to previously inaccessible riffle habitat that is used for spawning at this time of year (Gido and Propst, 2012; Gido et al., 2013). The positive relationship to higher numbers of low flow days in summer is likely related to the warmer temperatures of such conditions that enable rapid growth of larvae as well as more frequent reproductive bouts for this repeat spawner (Moore and Thorp, 2008; Gido and Propst, 2012). Finally, the negative association with high flow days in autumn may be due to the cooling effect of such late season events which also deposit fine sediment over spawning habitat thereby, reducing spawning and recruitment success (Gido and Propst, 2012).

The hydrologic metrics, which were found to be statistically significant in earlier work (Gido and Propst, 2012), were chosen because they quantified a facet of the flow regime that the functional predictor could not account for, i.e. the frequency of a type of flow day, as opposed to the timing, magnitude or duration of a flow event. In this instance, the explanatory power of the model \( r^2 = 0.67 \) is almost double that of the model fit by Gido and Propst (2012) \( r^2 = 0.35 \) and was able to identify new flow-ecology relationships for this species that were not previously detected with this dataset.

**The use and interpretation of functional linear models**

Functional linear models offer a novel approach to quantifying flow-ecology relationships that is conceptually appealing and complements the existing approach of using hydrologic metrics. Our models were successful in quantifying the relationships between fish species’ densities and community composition and the timing, magnitude and the duration of higher flows. In using a mathematical decomposition of the actual discharge, a functional approach removes the uncertainty that can be associated with calculating hydrologic metrics from flow records of different lengths (e.g. Kennard et al., 2010). However, the functional predictor variable, and therefore the regression coefficient, does not readily quantify the influence of variability of flows, though it is possible to quantify relationships to the rate of change of flow by taking the first derivative of the functional predictor variable (see Ramsay and Silverman, 2006). We illustrated the flexibility of the approach in the red shiner model by using hydrologic metrics that described a different facet of the flow regime from a functional predictor, for summer and autumn flows in tandem with a functional predictor for spring flows. This also demonstrates the obvious point that a purely functional approach will not always yield better results – a statistical model can only detect relationships that exist in the data. Nonetheless, using scalar predictor variables along with a functional predictor variable provides an avenue to incorporate hydrologic metrics that describe flow variability; an important hydrological driver of diversity (Poff and Allan, 1995; Puckridge et al., 1998; Biggs et al., 2005). It would also be plausible to include scalar habitat descriptors of water quality (i.e., temperature) and habitat quality (i.e., substrate composition) in future analyses. Decisions about variable selection should ideally involve some expert knowledge in tandem with information theoretic indices (such as cross-validation or AIC) to ensure models have biological relevance. As with multiple regression models, the coefficients in a model combining scalar and functional predictors quantify the effect of each predictor variable having controlled for the others in the model (Ramsay et al., 2006).

There are at least two clear applications where a functional approach may be useful for river ecology and management. The first is to identify important flow-ecology relationships in
relatively unstudied species where hypotheses about specific hydrologic metrics may not exist. The functional linear model can be used to identify time periods when higher flows have a significant relationship to such a species. If required, this information can be used to guide further analyses using hydrologic metrics calculated over specific intervals. The second application would be to use functional linear models to develop and test environmental flow recommendations. The predictive analysis in the second case study demonstrated how it is possible to test a set environmental flow scenarios directly by applying a fitted functional model to each. Our results showed concordance between the fitted regression coefficients and predicted outcomes under the environmental flow scenarios. Predicted densities for each species were above and below the 18-year average when flows were released at times of the year that coincided with positive and negative relationships to higher flows for each species respectively. It is important to note, however, that the lack of a statistically significant regression model for flannelmouth sucker suggests that predictions for this species should be interpreted with caution. Nonetheless, the results of an analysis such as this can be used to determine the preferred environmental flow scenario given the specifics of the management objectives. We identified a flow prescription (i.e. flow regime) that was predicted to result in a lower than average density of the invasive species, red shiner, and a greater than average density of the two native species, flannelmouth sucker and speckled dace. When the responses of multiple species are important a formal decision analysis may still be required to decide which scenario would be preferable, given the predicted outcomes of each species. However, the use of functional models provides a methodology to link the likely response of organisms of interest to potential environmental flows.

In addition to the functional linear models presented here, there are other functional approaches that may also be of use in prescribing environmental flows including models for functional response variables, the use of wavelets, Fourier series and harmonic analyses. An ecohydrological model for a functional response variable is one where the predictor variable may be river discharge and the response variable would be another time series such as habitat volume defined by water velocity and depth (Ramsay and Silverman, 2006). Such a model could be used to make real-time predictions of available habitat (e.g. Stewart et al., 2005). Wavelets are a statistical tool that can be used to quantify multi-scaled variability in time series data such as river flow or stream temperature that may arise due to daily, seasonal or annual periodicities (Steel and Lange, 2007). Fourier analysis has also been used to decompose the frequency of periodic fluctuations within the hydrograph and quantify metrics of flow variability, autocorrelation and the frequency, magnitude and timing of key flow events (Sabo and Post, 2008). These latter two approaches essentially provide alternative methods to generating ecologically relevant hydrologic metrics based on a functional transformation of the hydrograph. Each of these approaches has its own strengths and weaknesses for quantifying different elements of the flow regime and the choice among them will be naturally informed by the goal of the analysis (Table 2).

Although functional approaches, and in particular functional linear models, have the conceptual appeal of using the actual hydrograph as a predictor, there are obviously challenges associated with their use. A familiar challenge for researchers using hydrologic metrics is to select a subset that describes some facet(s) of the flow regime that is hypothesised to influence the response variable. There are statistical and ecological approaches for this (e.g., Olden and Poff, 2003; Jowett and Biggs, 2009; Poff et al., 2010) and important guidelines to minimize metric uncertainty (Kennard et al., 2010); however, the selection of metrics will ultimately
influence the results of any ecohydrological analysis. The challenges in fitting a functional model revolve around the nature of the predictor variable and the subsequent regression coefficient(s). The selection of the temporal window for the functional predictor variable has the potential to influence the results of the analysis (e.g. Ainsworth et al., 2011). As with standard regression analysis, the regression coefficients may change with the inclusion of additional predictor variables. Also, the type of functional transformation used will influence the analysis. For example, a Fourier transformation may be more appropriate to capture the seasonality of a multi-year predictor variable than the B-spline approach we used in this study. These may be new challenges for researchers and managers using functional linear models, however, they can be overcome by invoking ecological intuition.

We have shown that functional linear models have great promise in quantifying relationships between ecological response variables and river discharge. Functional data analysis is an area of active research in the statistical literature (e.g. Ainsworth et al., 2011) and there are on-going developments including, generalised functional linear models that take advantage of different error distributions (Müller and Stadtmüller, 2005), and Bayesian functional regression (Crainiceanu and Goldsmith, 2010). Developments such as these will enable ecologists and river managers to use different types of response variables, such as species’ occurrences as well as incorporating issues of scale and different types of predictor variables, such as microhabitat variables into these models. We encourage ecologists and river managers to investigate further this suite of statistical tools for use in ecohydrology and when developing standards for environmental flows.

Acknowledgements We thank Meryl Mims for useful discussions and assistance with the data for the first case study. We also thank Giles Hooker for helpful discussions on functional data analysis. Dr. Angus Webb and one anonymous reviewer also provided helpful comments which improved the manuscript. JDO was supported by the H. Mason Keeler Endowed Professorship and the Department of Defense - Strategic Environmental Research and Development Program (RC-1724).

REFERENCES


Table 1 Predicted densities of each species (count/m$^{-2}$) under five different environmental flow scenarios (shown in Fig. 4) to release the approximately 370, 000 ML available for this purpose in the San Juan River (Holden, 1999). Each flow scenario is defined by the timing and length of the environmental flows with 30, 60 and 90 day flows comprising mean daily flow rates of 143 m$^{-3}$s$^{-1}$, 71 m$^{-3}$s$^{-1}$ and 48 m$^{-3}$s respectively. The mean and standard deviation of fish density over the 18 years of sampling is also included for comparison.

<table>
<thead>
<tr>
<th>Hydrograph</th>
<th>speckled dace</th>
<th>flannelmouth sucker</th>
<th>red shiner</th>
</tr>
</thead>
<tbody>
<tr>
<td>No e-flow</td>
<td>0.183</td>
<td>0.034</td>
<td>0.133</td>
</tr>
<tr>
<td>Early spring 60 days</td>
<td>0.256</td>
<td>0.055</td>
<td>0.137</td>
</tr>
<tr>
<td>Late spring 30 days</td>
<td>0.235</td>
<td>0.036</td>
<td>0.480</td>
</tr>
<tr>
<td>Late autumn 30 days</td>
<td>0.067</td>
<td>0.009</td>
<td>0.002</td>
</tr>
<tr>
<td>Summer 90 days</td>
<td>0.263</td>
<td>0.026</td>
<td>0.11</td>
</tr>
<tr>
<td>Mean density/m$^{-2}$</td>
<td>0.311</td>
<td>0.039</td>
<td>2.745</td>
</tr>
<tr>
<td>(standard deviation)</td>
<td>(0.243)</td>
<td>(0.04)</td>
<td>(5.672)</td>
</tr>
</tbody>
</table>
Table 2: The relative suitability of four different approaches - hydrologic metrics, functional linear models, wavelets, Fourier transforms and harmonic analyses and - to characterizing different components of the flow regime. Values are informed by the work of Bradshaw and Spies (1992), Ramsay and Silverman (2006), Sabo and Post (2008) and the authors’ personal experience.

<table>
<thead>
<tr>
<th>Elements of the flow regime</th>
<th>Hydrologic metrics</th>
<th>Functional linear model</th>
<th>Wavelets</th>
<th>Fourier transforms and harmonic analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional descriptions of flow</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Functional predictor variable</td>
<td>Unsuitable</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Temporal autocorrelation</td>
<td>Moderate</td>
<td>Moderate</td>
<td>Moderate</td>
<td>High</td>
</tr>
<tr>
<td>Multi-scaled variability</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Key facets of the flow regime</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnitude</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Duration</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Moderate</td>
</tr>
<tr>
<td>Frequency</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Timing</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Rate of change</td>
<td>High</td>
<td>Unsuitable</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Variability</td>
<td>High</td>
<td>Unsuitable</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>
Fig. 1 A comparison of observed discharge and four key summary statistics over a twelve month period for two different systems in the USA. The two very different hydrographs have very similar hydrologic metrics over the interval shown.

<table>
<thead>
<tr>
<th></th>
<th>WR</th>
<th>DC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean daily flow (m$^3$/s$^{-1}$)</td>
<td>15.6</td>
<td>16.2</td>
</tr>
<tr>
<td>Constancy</td>
<td>0.64</td>
<td>0.71</td>
</tr>
<tr>
<td>Rate of rise (m$^3$/s$^{-1}$/d$^{-1}$)</td>
<td>0.28</td>
<td>0.36</td>
</tr>
<tr>
<td>Annual CV</td>
<td>1.57</td>
<td>1.86</td>
</tr>
</tbody>
</table>
**Fig. 2** Map of the USA showing the location of sampling sites for the two case studies. The first case study comprised a database of fish species occurrences from one-time standardised surveys at 32 free flowing streams across Midwestern USA. The second case study consists of a single sample of the density of three different species collected each year over 18 years from secondary channels in a single 138 km long reach of the San Juan River in New Mexico and Utah from 1993-2010.
Fig. 3 Observed discharge in the San Juan River (log scale) for two different years (thick dashed line) with the functional predictor variable used for each year overlaid (thin dotted line).
Fig. 4 Five hypothetical environmental flow scenarios using the approximately 370,000 ML of water available for such use in the Navajo Dam. In each scenario the entire volume of water is delivered as part of a single flow release over 30, 60 or 90 days which equates to releases at mean daily flow rates of 143 m$^3$s$^{-1}$, 71 m$^3$s$^{-1}$ and 48 m$^3$s$^{-1}$ respectively.
Fig. 5 Functional regression coefficients for the proportion of the three life history strategies; periodic, opportunistic and equilibrium, in streams across the Midwestern USA. The functional coefficients quantify the relationship between the richness of each life history strategy and river flow on each day of the year prior to sampling. Streams with higher runoff at times of the year when the coefficient is above zero tend to have higher proportions of each strategy during the summer sampling period. In contrast, streams with higher runoff when the coefficient is below zero tend to have lower proportions of each strategy during the summer sampling period. The solid line represents the parameter estimate with dotted lines being the 95% confidence interval. The p-values on each plot indicate the significance of that model based on the permutation tests described in the Methods.
Fig. 6 Functional regression coefficients for the density of speckled dace, flannelmouth sucker and red shiner in the San Juan River. See caption for Fig. 4 for an explanation of the functional coefficient.
Fig. 7 Functional and scalar regression coefficients for the second model fitted for red shiner. This model provided a much better fit to the data than the first model for red shiner (Figure 5c). As with the functional coefficient, the scalar regression coefficients are shown with their approximate 95% confidence intervals, neither of which cross zero, indicating a significant relationship. The two scalar predictors are number of days in summer <14 m$^3$s$^{-1}$ (LFS) and the number of days over the 75th percentile of flow in autumn (HFA).